

Ph.D. Research Proposal

Doctoral Program in “Department Name”

User Level Security Management and Insider Threat
Detection from Cyber Activities using CNN-LSTM

Model



by

<Name of the Candidate>

<Reg. No of the Candidate>

<Supervisor Name>

<Date of Submission (DD MM 20YY)>

I. INTRODUCTION / BACKGROUND

Over the past few decades, insider threat is one of the most hazardous and prevalent security threats to much organization, institutes, government entities and more. Since, insider threats occur by authorized worker within the organization. Hence, detecting the insider threat in organization is one of big issue and hardest to identify [1]. Thus induce many researchers focused towards this area to introduce new approach to detect the insider threat in organizations. Insider threats are modeled as those who steal the organization's information, commit work place violence, etc.

The insiders can be employees, partners, or ex-employee. The intention of insider is also reflected in his/her social networks [2]. Further, the insider threats are the one of the significant reasons for phishing susceptibility [3]. Thus it is necessary to mitigate the insider threats in order to prevent the organization from many threats. Behavior analysis is considered as one of the optimal factors for insider threat detection [4]. In addition to behavior analysis, sentiment analysis is also concentrated by some researchers [5].

The user behavior modelling based insider threat detection process is executed in [6]. Here, the four different anomaly detection algorithms are utilized to detect the insider threat. The user behavior related features such as user daily activity, email contents and user communication history are considered to perform the insider threat detection process. In study [7], the scenario based insider threat detection process is utilized. Herein, the two layered auto encoder model is utilized to detect the insider threat. . A deep neural network (convolutional neural network) and long short term memory (LSTM) are used to model the insider behaviors [13].

A one-class naïve Bayes algorithm is used to identify the mouse dynamics for insider threat detection [14]. However, naïve Bayes algorithm fails when the features are dependent. Overall, following issues are still exist in insider threat detection,

- Random feature selection increases detection time and minimizes the accuracy

- Insider threat detection must consider the behavior as well as sentiment analysis for accurate threat detection

1.1 Research outline & Scope

The main aim and scope is to detect the insider threat with high accuracy rate using both sentiment and behavioural information. In addition, it is also aim to overwhelm issues faced in previous works during insider threat detection.

1.1 Research Objectives

The major objective of this work is to provide user level security and semantic meaning based insider threat detection. The key objectives of proposed work are listed as follows,

- To enhance the dataset contents into rich content using the effective preprocessing step. In order to reduce the difficulties in insider threat detection processes.
- To increase the detection accuracy using both user behaviour and sentiment related scores.
- To fasten the feature extraction process use of MapReduce in order to reduce the computation time of insider threat detection process.
- To construct email content ontology in order to extract sentiment score effectually from the dataset.

To reduce the fast feature set during the insider threat detection process using optimization based feature selection process.

II. RESEARCH GAPS

2.1 Common Problem Statement

So far, there have been many issues remains during the insider threat detection. They are discussed as follows: The malicious activity of the insider is not occurred frequently that leads to unavailability of the data required for threat detection. Hence, in-depth details of user behavior

model must be considered during insider threat detection. Further handling of huge amount of data also introduced many issues. Since, organization environment have numerous amount of data such as File Access Logs, Email History, Network Traffic and more. These issues induce many difficulties during insider threat detection.

2.2 Problem Definition

In [16] data granularity levels are utilized to perform insider threat detection. Here, the machine learning algorithm is utilized. This paper utilized CERT r5.2 dataset for insider threat detection process. At first, it performs preprocessing process where data gathered from the different source of organization are aggregated. In this, data aggregation is performed based on the time duration and actions performed. And then, features are extracted such as HTTP, Email, File, USB and login. After that, extracted features are classified using the four different machine learning algorithms such as RF, Logistic Regression, Neural Network and XGBoost.

Problem

- During pre-processing process, data aggregation step handled that cannot enhance the data present in the collected database. Since, CERT dataset has more various unit data, missing values, redundant data, combined representation of year and time. Thus requires most significant preprocessing step afore to the detection process.
- High specific features (LDAP, psychometric, web) are not extracted during the feature extraction process. Hence, detecting insider threat becomes tedious that tends to reduce the detection accuracy.
- Here, features are not extracted using effective framework thus increases the time required to detect insider threat. Since, CERT has more amount of data with its features thus takes more time to complete the insider threat detection process.

Proposed

- In our work, we perform different processes in preprocessing step that are data cleaning, data parsing, data reduction and standardization. These processes enhance the content in the database.
- In our work, we extract high significant features such as web and file access, email, HTTP, USB connection, Sentiment, LDAP, decoy file, psychometric and private email utilization. These features improve the insider threat detection process.
- In our work, we utilize effective CNN framework and deep learning algorithm to detect insider threat. Thus avoids the high processing time and tedious processing in insider threat detection.

In [17] paper uses EEG signal for user authentication and access control. Here the insider threat detection is performed based on the physiological measurements of the intent based access control model. Herein the behavior of a user is validated by the micromovements captured by the EEG data. The major objective of this work is to implement intention based insider threat detection.

Problems

- Although intention based solution is better, it also needs to consider the behavior and mouse dynamics of the user to detect insider threat
- In this work, authenticating user through EEG signal has lower practical possibility and needs high cost to implement (requires sensors, headphones etc.)
- Not suitable for continuous authentication (however, intention can be changed at any time)

Proposed Solutions

- Mouse dynamics are considered in FP_AUTH based authentication scheme along with iris based pre-authentication which is practical
- Mouse dynamics are extracted for continuous authentication

The authors of this [18] paper have proposed a hybrid model of Across-Domain Anomaly Detection (ADAD) and Across-Time Anomaly Detection (ATAD) based on isolation forest algorithm and improved Markov model respectively. According to the unusual behavior of the user the scores are obtained from individual model and then they are fused into one for each user. Domain features (logon events, logoff event, removable media and file copy events) are selected using Principal Component analysis (PCA).

Problems –

- PCA is not supported for non-linear data and hence it results with poor results in insider threat detection
- In Markov chain the status of user is determined by taking in account of recent history of the particular user. However, it is essential to consider past behavior of users for accurate insider detection.

Proposed Solutions

- The proposed work supports with the processing of linear data
- Multiple features are taken in account along with the consideration of history score for evaluating his/her past behavior.

This paper proposes [19] the simultaneous neural learning method to perform insider threat detection process. This study performs insider threat detection on CERT r6.2 dataset. In this data from the security log is considered for the insider detection. Initially, it transforms the dataset content to the text using the log2text model. And, further converted text is transformed into the corpus with the aid of the text2corpus model. The generated corpus is trained using the word2vec model. Finally, bayes probability model is executed to detect the insider threat.

Problem

- Lack of preprocessing step afore to the insider threat detection induces difficulties during insider threat detection. Since, CERT dataset has more redundant data, missing value,

multi-unit data and more. These difficulties result in lower accuracy in insider threat detection.

- Here, all features are extracted and converted to the vector form using the skipgram model which produce high dimension feature vector. Thus creates difficulties during the insider threat detection process.
- Here, bayes probability is used to detect the insider threat which consumes more time. In addition, it produces accurate result when the vector size is small. As a result, detection rate is reduced.

Proposed

- In our work, we have performed different preprocessing processes such as data cleaning, data reduction, parsing and standardization to easier the further insider threat detection process.
- Our work extracts feature using CNNframework thus doesn't generate high feature vector thus doesn't introduce difficulties during insider threat detection process.

In this [20] paper, an internal threat detection model is designed. The authors have highlighted that the unsupervised learning algorithm is better than supervised learning algorithm. Denoising autoencoder is proposed for insider threat detection. The denoising autoencoder is utilized for feature extraction. Finally, the Gaussian mixture model is presented for anomaly detection. To find best Gaussian mixture model, a cross-validation method is proposed.

Problems

- This work uses robust invariance, OCSVM, isolation-forest, and local outlier factor are considered for anomaly detection which increases time consumption and complexity
- In addition, parameter tuning by cross-validation method is not efficient which affects the accuracy of classification

Proposed Solutions

- Integrated sentiment and behavior analysis is proposed with CNN

III. RESEARCH CONTRIBUTIONS

To overwhelm above mentioned problems, we propose a novel Tri Steps based Insider Threat Detection (T3ITD) Model driven by Behavior and Sentiment Analysis. Our proposed model consists of following modules for insider threat detection,

i) Strong Authentication

Initially, all users (i.e.) employees of the organization are authenticated to access the system. Proposed authentication module is function upon two-folds as (i) Biometric based pre-authentication, and (ii) Mouse Dynamics based continuous authentication. In first stage, the user is authenticated based on {ID, PW, BM}. Here the considered biometric (BM) is iris of the users. Then the pre-authenticated users are authenticated based on the mouse dynamics. For continuous authentication, we proposed a Frequent Pattern based Authentication (FP_AUTH) scheme in which the past behaviors of that user is clustered based on frequent patterns. Further authentication weight (Auth_Weight) is formulated by computing the Hassanat Distance between current and previous pattern of that user.

ii) Bi-Examination

As we mentioned earlier, many of the existing works are focused on either behavior analysis or sentiment analysis. We proposed an integrated assessment module for insider threat detection. We analyze the user behavior from activity log and sentiment from email content. At first Data Augmentation is performed to improve the quality of the data. This step includes, *Data Cleaning, Data Transformation using BoxCox Transformation, and Data Reduction using Isomap algorithm*. For that we propose a novel Convolutional Neural Network (CNN) which performs both sentiment and behavior analysis for each user. In CNN, the first layer pre-processes the data for improving the detection accuracy. Then, optimal features are selected by GrassHopper Optimization (GHO) algorithm to reduce the feature dimensionality. In sentiment analysis, we consider optimal features such as polarity score, attachment, mail ID, and time stamp. Finally, the CNN provides the *behavior score (Behave_Weight) and sentiment score (Senti_Weight)*.

iii) Intelligent Decision Making

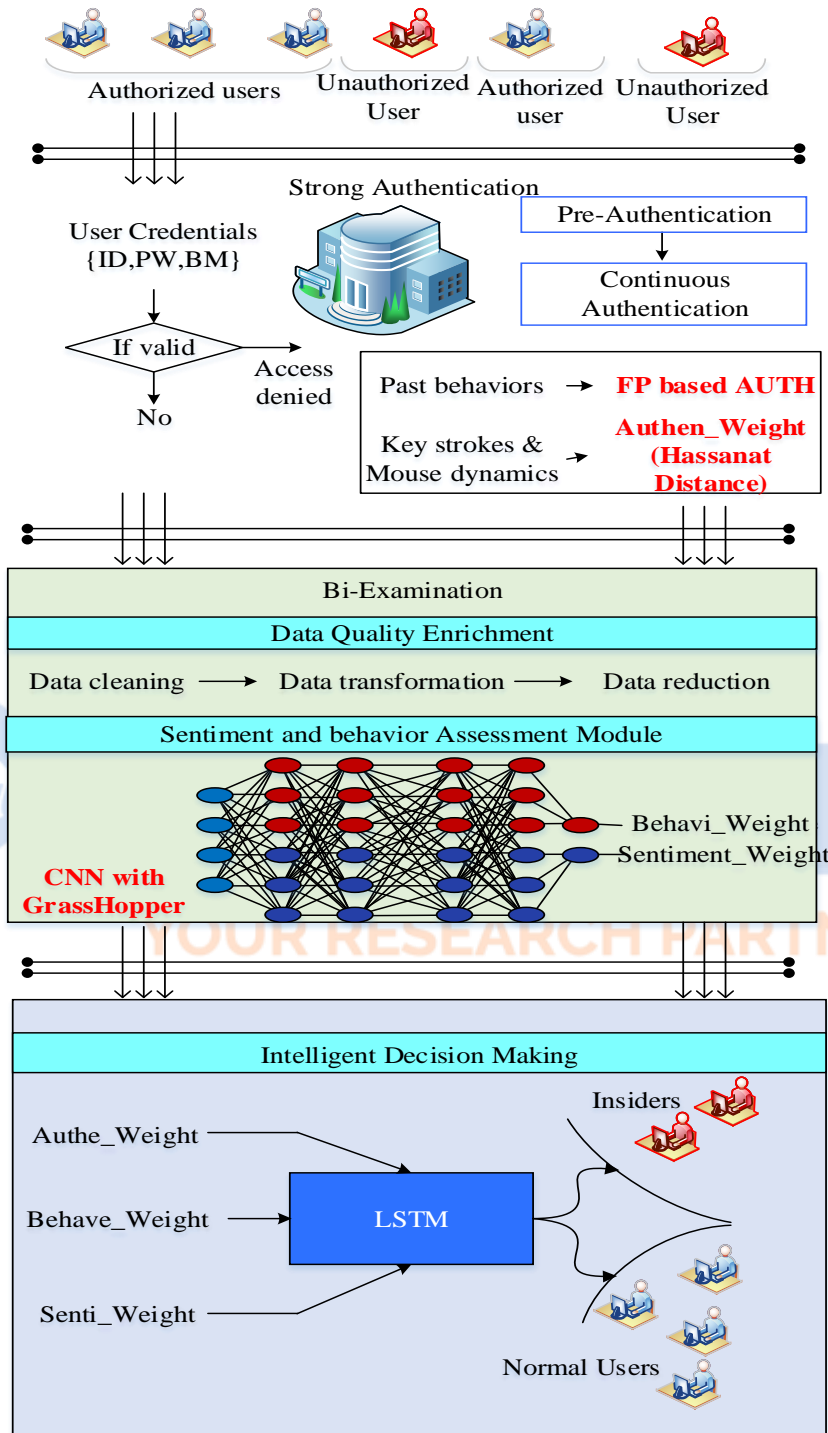
This module is responsible for making decision on insider. For each user, the Auth_Score, Behave_Score, and Senti_Score are computed from prior modules. These scores are fed as input of **LSTM** and the insider threat is detected based on the LSTM output. Thus considering, mouse dynamics, behavior, and sentiment improves the accuracy of insider threat detection. This work considers CERT 5.2 dataset for our insider threat detection process.

Performance Evaluation

Finally, our work is evaluated based on following performance metrics,

- Precision
- Recall
- F-Score
- Accuracy
- ROC
- False Positive Rate
- Computation Time





IV. RESEARCH NOVELTIES

The highlights of the proposed work are discussed as follows:

- We executed four different steps in the preprocessing phase in order to enhance the database content such as data cleaning, data reduction, data parsing and standardization
- We utilized CNN to extract the features from the dataset and also estimate the weight score for each user behavioural features using the CNN.
- We select the optimal features from the extracted feature using GHO algorithm in order to reduce the feature dimension that tends to reduce the insider threat detection time.
- Our work constructs sentiment analysis using the email content and further utilizes to estimate the sentiment score.
- Finally, we execute LSTM algorithm to detect the insider threat which utilize both weight score and sentiment score.

V. PREVIOUS WORKS & LIMITATIONS

Paper 1

Title – An Adversarial Risk Analysis Framework for Cybersecurity

Concept –

This paper is presented so mitigate the risk of both intentional threats and non-intentional threats. Insiders are the intentional adversaries those who participate in an organization. By the defined security controls, the attacks are detected. This paper majorly highlights the risk and severity of detecting threats that are participating in the organization.

Paper 2

Title – Insider Threat Detection with deep Neural Network

Concept –

In this paper the authors have proposed long short term memory (LSTM) with convolution neural network (CNN) for evaluating the user behavior. LSTM extracts the features and then the insiders are detected in CNN. The dataset used in this work is CERT 4.2 for detecting the insiders. The CNN utilizes fixed size of features for detecting the insiders.

Paper 3

Title: Improving Insider Threat Detection through Multi-Modelling/Data Fusion

Concept

This paper deals with the insider threat detection through multi-modeling and data fusion techniques. Insider threats are very harmful for any organization and can be modeled in various forms. Some general insider threats include those who steal the data, commit workplace violence, or commit any other act that is detrimental to the organization. The insider can be business partners, employees, or others who are trying to expose the private data of the organization.

Paper 4

Title: A New Approach to Modelling the Effects of Cognitive Processing and Threat Detection on Phishing Susceptibility

Concept

This paper designs a new approach for insider threat detection with cognitive processing. This paper analyzes the insider threat in the perspective of phishing susceptibility. The attack is analyzed in terms of attack quality, motivation to process, ability to process, and knowledge. Insider threat detection is determined as the major solution to mitigate the phishing susceptibility. In other words, involvement of insider threats leads to many other attacks like phishing.

Paper 5

Title: User Profiling in Anomaly Detection of Authorization Logs

Concept

This paper presents a user profiling method for anomaly detection inside the organization. In addition, the analysis of log authorization method is also proposed to improve the efficiency. This method allows the companies to assess each user's activities and detect the even slight variation from the usual pattern.

Paper 6

Title: Implementation of Insider Threat Detection System Using Honeypot Based Sensors and Threat Analytics

Concept

In this paper, the authors have highlighted that the current insider threat detection methods are not suitable for real-time applications. Then a encrypted honeypots based insider threat detection mechanism is proposed. In general, the insiders are capable to violate the security approach or entrance control of an organization. The honeypot sensor is used to acquire the information from the virtual machines.

Paper 7

Title: One-class naive Bayes with duration feature ranking for accurate user authentication using keystroke dynamics

Concept

This paper presents a feature ranking method for user authentication through mouse dynamics analysis. In this work, the speed on specific keys and the key stroke's index order are considered as the optimal features and the users are authenticated based on them. The machine learning algorithm namely one-class naïve Bayes algorithm is incorporated to detect the unauthorized users through mouse dynamics.

Limitations

- Naïve Bayes algorithm has shortcoming that it fails when the attributes are dependent
- Another biometric can be utilized in addition to the mouse dynamics for strong authentication

Paper 8

Title: Insider Threat Detection: Machine Learning Way

Concept

This paper uses machine learning approach for insider threat detection. In this work, USB device insertion and removal event is mainly focused for insider threat detection. The malicious activities are identified by applying linear regression and Cook's and Mahalanobis distance measure. In addition, neural network and support vector machine algorithms are applied for login activities to successfully demonstrate the detection of anomaly behavior.

Limitations

- This work only focuses on USB drive activities. However, it is necessary to consider the other constraints for accurate insider threat detection.
- In general, the dataset involves multiple features thus it is required to select optimal feature subset to optimally detect the insider threat

Paper 9

Title: An Insider Threat Detection Method Based on User Behavior Analysis

Concept

This paper proposes a behavior analysis model for insider threat detection. At first, the user behaviors are aggregated over a time period and then multi-domain features are extracted from the audit log. Finally, the XGBoost algorithm is proposed to train the data for detecting the insider threats. The audit log involves the system logon/logoff, file access, device usage, HTTP access, mail sending, and receiving records.

Limitations

- This method uses random features extracted from the audit logs which is not efficient
- Initial authentication is required to prevent the non-employee access in the organization

Paper 10

Title: Keystroke dynamics-based user authentication using freely typed text based on user-adaptive feature extraction and novelty detection

Concept

Password based authentication methods are not feasible and have more possibility to be cracked by the attackers. The better solution for this issue is biometric based authentication. However, biometric is also forged by many cases which reduces the strength of the authentication scheme. To overcome both issues, this paper uses the keystrokes and mouse dynamics to authenticate the users continuously.

Paper 11

Title – AnyThreat: An Opportunistic Knowledge Discovery Approach to Insider Threat Detection

Concept –

This paper proposes an opportunistic knowledge discovery system for detecting insiders in the system. This AnyThreat consists of four following components as (1) feature engineering, (2) oversampling, (3) class decomposition and (4) classification component. From the samples, the positive and negative nearest neighbors are determined using Synthetic Minority Oversampling Technique (SMOTE). Based on the majority and minority classes the users are classified.

Paper 12

Title- Detecting Potential Insider Threat: Analyzing Insiders Sentiment Exposed in Social Media

Concept

In this paper, insider threat is detected using the sentiment exposed in social media model. For insider threat detection, it utilizes the machine learning algorithm namely decision tree and K-Means algorithms. Here, the user features are extracted using topic modelling technique. Here, the extracted features are processed with the aid of the machine learning algorithms both supervised and unsupervised algorithms. It concludes decision tree and k-means algorithms provide better result in insider threat detection.

Limitations

- The optimal feature selection process is required since there have been many redundant data present in the extracted features. Here feature selection process is not handled thus increases difficulty during insider threat detection process.

Paper 13

Title- A Fine-grained Approach for Anomaly Detection in File System Accesses with Enhanced Temporal User Profiles

Concept

In this paper, anomaly detection is executed using the fine grained approach on four synthetic datasets. Here, the fine grained activity of the user is considered based on the time manner to detect the insider threat. In this study, user profile is created based on three different model including block level profiler, frequency profiler and access cluster profiler. Besides, it also performs frequency profiling process to perform insider threat detection. Based on the collected temporal user profile, this study classifies the anomaly detection in the given dataset.

Paper 14

Title: Scenario-Based Insider Threat Detection From Cyber Activities

Concept

This paper presents a scenario based insider threat detection technique for cyber activities. This work uses time-series classification of user-activities for threat detection. Single-day features are computed from the activity logs. Further, autoencoder is applied to classify the normal and anomaly data for insider threats. In this work, CMU-CERT r.4.2 dataset is used for validation.

Limitations

- CMU-CERT dataset is a synthetic dataset which contains noises and has high redundant data. Thus it is necessary to preprocess the before classification reduces the classification accuracy
- Autoencoder has limitation in testing the data (it classifies the data accurately when the test data is more likely to training data but in practical the insider data is varying)
- This work ignores the psychometric data in threat detection which is also significant for insider threat detection

Paper 15

Title- Role-based Log Analysis Applying Deep Learning for Insider Threat Detection

Concept

This paper utilized the role based log analysis model to perform insider threat detection process. In this, CERT r4.2 dataset is utilized to extract the features in order to detect the insider threat. Here, the deep learning algorithm is used to classify the insider threats in the given dataset. For this purpose, it proposes long short term memory algorithm which learns the feature from the user different daily behaviours. Based on the learned features from the user role based log model, this study classifies the insider threat using the long short term memory algorithm.

BIBLIOGRAPHY

David Rios Insua, Aitor Couce-Vieira, Jose A. Rubio, Wolter Pieters, Katsiaryna Labunets, Daniel G. Rasines, “An Adversarial Risk Analysis Framework for Cybersecurity”, Risk Analysis, Online Wiley, 2019.

- Fangfang Yuan, Yanan Cao, Yanmin Shang, Yanbing Liu, Jianlong Tan, Binxing Fang, “Insider Threat detection with Deep Neural Network”, Computational Science, Springer, pp. 43 – 54, 2018.
- Brown, D.E., Buede, D.M., & Vermillion, S.D. (2019). Improving Insider Threat Detection Through Multi-Modelling/Data Fusion. *Procedia Computer Science*, 153, 100-107.
- Musuva, P.M., Getao, K.W., & Chepken, C.K. (2019). A new approach to modelling the effects of cognitive processing and threat detection on phishing susceptibility. *Computers in Human Behavior*, 94, 154-175.
- Zamanian, Z., Feizollah, A., Anuar, N.B., Kiah, L.B., Srikanth, K., & Kumar, S.S. (2019). User Profiling in Anomaly Detection of Authorization Logs. *Computational Science and Technology*, 59-65.
- Yamin, M.M., Katt, B., Sattar, K., & Ahmad, M.B. (2019). Implementation of Insider Threat Detection System Using Honeypot Based Sensors and Threat Analytics. *Lecture Notes in Networks and Systems*, 70.
- Ho, J., & Kang, D. (2017). One-class naïve Bayes with duration feature ranking for accurate user authentication using keystroke dynamics. *Applied Intelligence*, 48, 1547-1564.
- Raval, M.S., Gandhi, R., & Chaudhary, S. (2018). Insider Threat Detection: Machine Learning Way. *Versatile Cybersecurity*, 19-53.
- Jiang, W., Tian, Y., Liu, W., & Liu, W. (2018). An Insider Threat Detection Method Based on User Behavior Analysis. *Intelligent Information Processing*. 421-429.
- Kim, J., Kim, H., & Kang, P. (2018). Keystroke dynamics-based user authentication using freely typed text based on user-adaptive feature extraction and novelty detection. *Appl. Soft Comput.*, 62, 1077-1087.
- Diana Haidar, Mohamed Medhat Gaber, Yevgeniya Kovalchuk, “AnyThreat: An Opportunistic Knowledge Discovery Approach to Insider Threat Detection”, arXiv, 2018.

- Park, W., You, Y., & Lee, K. (2018). Detecting Potential Insider Threat: Analyzing Insiders' Sentiment Exposed in Social Media. *Security and Communication Networks*, 2018, 7243296:1-7243296:8.
- Mehnaz, S., & Bertino, E. (2019). A Fine-grained Approach for Anomaly Detection in File System Accesses with Enhanced Temporal User Profiles. *IEEE Transactions on Dependable and Secure Computing*, 1-1.
- Chattopadhyay, P., Wang, L., & Tan, Y. (2018). Scenario-Based Insider Threat Detection From Cyber Activities. *IEEE Transactions on Computational Social Systems*, 5, 660-675.
- Zhang, D., Zheng, Y., Wen, Y., Xu, Y., Wang, J., Yu, Y., & Meng, D. (2018). Role-based Log Analysis Applying Deep Learning for Insider Threat Detection. *SecArch'18*.
- Le, D.C., Zincir-Heywood, N., & Heywood, M.I. (2020). Analyzing Data Granularity Levels for Insider Threat Detection Using Machine Learning. *IEEE Transactions on Network and Service Management*, 17, 30-44.
- Almehmadi, A. (2018). Micromovement Behavior as an Intention Detection Measurement for Preventing Insider Threats. *IEEE Access*, 6, 40626-40637.
- Bin Lv, Dan Wang, Yan Wang, Qiu Jian Lv, Dan Lu, "A Hybrid Model Based on Multi-dimensional Features for Insider Threat Detection", *Wireless Algorithms, Systems, and Applications*, pp. 333 – 444, 2018.
- Liu, L., Chen, C., Zhang, J., Vel, O.D., & Xiang, Y. (2019). Insider Threat Identification Using the Simultaneous Neural Learning of Multi-Source Logs. *IEEE Access*, 7, 183162-183176.
- Zhang, Z., Wang, S., & Lu, G. (2020). An Internal Threat Detection Model Based on Denoising Autoencoders. *Advances in Intelligent Information Hiding and Multimedia Signal Processing*, 391-400.